ARMSTRONG LABORATORY

AL/CF-TR-1995-0149

THE EFFECTS OF RESOLUTION AND OTHER FACTORS ON TARGET ACQUISITION PERFORMANCE USING SIMULATED SYNTHETIC APERTURE RADAR IMAGERY

Judi E. See

LOGICON TECHNICAL SERVICES, INC.
P.O. BOX 317258
DAYTON, OH 45431-7258

Gilbert G. Kuperman

CREW SYSTEMS DIRECTORATE
HUMAN ENGINEERING DIVISION
WRIGHT-PATTERSON AFB OHIO 45433-7022

JULY 1995

INTERIM REPORT FOR THE PERIOD OCTOBER 1994 TO JULY 1995

AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

TECHNICAL REVIEW AND APPROVAL

AL/CF-TR-1995-0149

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

William T. Wittman

for **KENNETH R. BOFF**, Chief
Human Engineering Division
Armstrong Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
July 1995

3. REPORT TYPE AND DATES COVERED
Interim Report Oct 1994 - Jul 1995

4. TITLE AND SUBTITLE
The Effects of Resolution and Other Factors on Target Acquisition Performance Using Simulated Synthetic Aperture Radar Imagery

5. FUNDING NUMBERS
C: F41624-94-D-6000
PE: 62202F
PR: 7184
TA: 10
WU: 44

6. AUTHOR(S)
*Judi E. See
Gilbert G. Kuperman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Logicon Technical Services, Inc.
P.O. Box 317258
Dayton, OH 45431-7258

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory, Crew Systems Directorate
Human Engineering Division
Human Systems Center
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7022

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL/CF-TR-1995-0149

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Synthetic Aperture Radar (SAR) sensors offer all-weather and stand-off range capabilities when employed against target acquisition tasks. SAR may be used by attack, surveillance, and/or reconnaissance platforms to detect, locate, track, and identify mobile targets such as theater ballistic missile launch vehicles. Operator performance with SAR varies as a function of sensor resolution, sensor employment geometry (grazing angle), and overall scene clutter. This study manipulated these variables through the generation of controlled simulated SAR imagery.

14. SUBJECT TERMS
Target Acquisition, Synthetic Aperture Radar (SAR), Mobile Missiles, Synthetic Imagery, Operator Performance

15. NUMBER OF PAGES
48

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

This page intentionally left blank.

ABSTRACT

Operator performance data were collected for simulated synthetic aperture radar (SAR) images that varied in grazing angle (15°, 30°, and 45°), resolution (8.5', 5', 3', or 1'), and background clutter (low, medium, and high). The performance outcomes that were analyzed included percentages of hits and false alarms, reaction time, and the signal detection theory measures of perceptual sensitivity (d') and response bias (c). Receiver operating characteristic (ROC) curves were also generated for each of the four levels of image resolution. Examination of the primary variable of interest in this study, image resolution, revealed that optimal performance in terms of operators' perceptual sensitivity occurred at the 3' resolution. This level of image resolution was also associated with faster reaction times for both correct detections of target objects and correct rejections of nontargets. Enhancing resolution further to the 1' level did not affect sensitivity to target detection but did yield faster reaction times for correct rejections. The obtained values of sensitivity for each resolution and their concomitant hit and false alarm proportions are to serve as inputs in an engagement-level model that utilizes sensor and operator performance data to perform mission effectiveness analyses on airborne systems engaged in target acquisitn.

PREFACE

This effort was conducted by the Crew Systems Integration Branch of the Armstrong Laboratory (AL/CFHI) under Work Unit 71841044, "Crew-Centered Aiding for Advanced Reconnaissance, Surveillance, and Target Acquisition." It was supported by Logicon Technical Services, Inc. (LTSI), Dayton, Ohio, under Contract F41624-94-D-6000, Delivery Order 0004. Lt Bryan Christensen was the Contract Monitor.

The authors wish to acknowledge the Air Force Theater Missile Defense Attack Operations Program Office of the Aeronautical Systems Center (ASC/LAAT) and of the Air Superiority Division of Headquarters, Air Combat Command (ACC/DRA) for support and encouragement.

Special recognition is due to the following LTSI personnel: Mr. Ken Crum who prepared the simulated imagery, Mr. Dave Robinow who created the software which controlled the operator performance experiment, and to Mrs. Iris Davis who performed the actual data collection.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page intentionally left blank.

INTRODUCTION

Background

The Air Force is conducting a Concept Definition and Exploration (CD&E) program in the Theater Missile Defense (TMD) Attack Operations (AO) mission sub-area. The purpose of the CD&E program is to improve the capability of the United States to detect, locate, track, identify, attack, and kill not only theater missile (TMs) systems but also their supporting command and control capabilities and associated infrastructure. Headquarters, Air Combat Command (ACC), Langley Air Force Base, Virginia, is directing a study to develop a quantitative analytical foundation to support decisions impacting the development, production, and fielding of TMD AO enhancements to find TMs, task surveillance and attack resources, and attack these targets. Sensor technologies (sensors, sensor management subsystems, automatic target cueing and recognition [ATC/ATR] subsystems, and operator interfaces) are expected to contribute significantly to achieving the required operational capabilities.

Synthetic aperture radar (SAR) sensors, on both surveillance and attack platforms, are of particular interest since this class of sensor is robust even under adverse weather conditions, can be employed over long standoff distances, supports accurate geolocation of detected/identified targets, and is capable of producing the level of image quality required for high confidence target acquisition and fratricide avoidance. Currently, SAR sensors are or will be employed on the J-STARS (E-8C) developmental surveillance system, the U-2R reconnaissance system, the reinstated SR-71 strategic reconnaissance system, several developmental unmanned air vehicle

(UAV) reconnaissance systems, and the F-15E, B-1B, and B-2 attack systems to support navigation and weapon delivery. Any or all of these sensor systems might be upgraded to support the TMD AO target acquisition mission.

Objective

Baseline operator performance data are required to support the operational effectiveness analysis of TMD AO concepts. The operators may be either a weapon system officer (WSO) or image analyst (IA). Prior research was generally concerned with the detection, recognition, and identification of tactical targets (e. g., tanks) or strategic relocatable targets (e. g., SS-25 ICBM systems). In general, the targets, backgrounds, flight profiles, and other major factors considered in the extant research are radically different from those associated with the TM problem. Imagery collection projects are underway with operational and developmental SAR assets, primarily to support ATC/ATR algorithm development, which will correct these limitations. Operator performance data are needed immediately, however, to support the analytic efforts currently in progress.

Since performance data do not yet exist and since content-valid imagery is not yet readily available, sensor imagery simulation offers an attractive alternative for collecting the required performance data. In the present study, simulated SAR imagery was used in a part-task target acquisition setting. High fidelity computer-aided design (CAD) models of SCUD TM transporter/erector/launchers (TEL) were embedded in generic backgrounds. These data bases were input to a SAR imagery generation software package to produce simulated imagery.

Resolution

The primary goal of the current study was to determine how SAR resolution, or the quality of the image produced by the sensor, affects operator performance. At very low resolutions, gross features such as mass and shape (i.e., "blobology") may be the only perceptible characteristics of a potential target. At higher resolutions, finer details of the object can be discerned. In general, as image resolution increases, two objects can occur in closer proximity to each other and still be perceived as separate objects rather than as a single entity. Previous investigations of the impact of image resolution on the detection of tactical or relocatable targets have indicated that performance accuracy improves with image resolution, though such improvements may become negligible as resolution increases beyond an already high level (Kuperman, Wilson, & Davis, 1993; Kuperman, Wilson, & Perez, 1988). In order to determine whether the same pattern extends to the detection of TMs, four levels of resolution ranging from very low (8.5') to very high (1') were examined in the present study. At the poorest resolution, two objects had to be at least 8.5 feet apart in order to be confidently distinguishable as two separate objects; whereas at the highest resolution, they could be separated by as little as 1 foot.

Background clutter

A secondary aim in the present study involved examining the effects of background clutter on performance effectiveness. Background clutter is generally viewed as the busyness of the scene in which a potential target object is embedded and may include both natural and manmade sources of clutter (Toms & Kuperman, 1991). High levels of background clutter may be characterized by the presence of geographical features such as dense forests that can make a target more difficult to discriminate or by the presence of large numbers of confusing objects (e.g., nontargets, decoys, and other target-like objects). Low levels of clutter might be presented by a desert scene with minimal vegetation or the presence of few confusing objects.

In an investigation of the effects of background clutter on the detection of relocatable targets, Kuperman, Wilson, and Perez (1988) defined clutter as the amount of vegetative coverage in a scene and specified four different types of clutter: FOREST (greatest overall vegetative coverage and tree height), TREES (moderate coverage and moderately tall trees), BUSHES (short trees and brush with only a few tall trees), and RIVER (a river with fairly sparse forest coverage). Performance efficiency was lowest in the FOREST condition and increased progressively as the background clutter decreased from TREES to BUSHES to RIVER. In the current investigation, clutter was characterized in a similar fashion as the number of trees per square mile in the scene.

Grazing angle

The final independent variable included in this study, grazing angle, refers to the angle formed between the ground and the line of sight from the sensor to the target, as depicted in Figure 1. Low grazing angles tend to result both in more pronounced terrain masking and in the occurrence of elongated shadows that can obscure any objects present in the scene. In addition, at low grazing angles, ridge lines and tree lines appear more intense, due to a radar phenomenon termed "front edge highlighting," which can further hinder target detection. Given these effects, one would expect low grazing angles to be associated with poorer performance than higher angles. To date, grazing angle per se has apparently not been included as an independent variable in target acquisition studies.

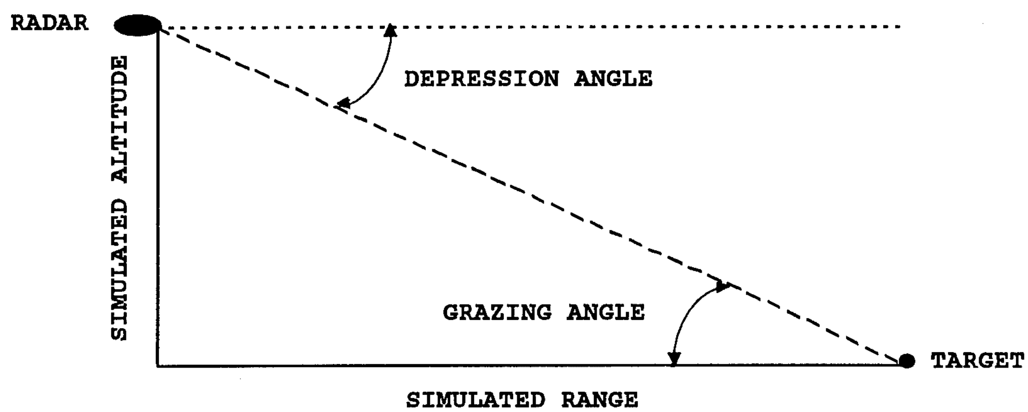

Figure 1. A geometric representation of grazing and depression angles.

However, as reported by Spravka, Crawford, and Kuperman (1990), several studies have looked at the effects of depression angle. As can be seen in Figure 1, depression angle is the angle formed between the local horizontal at the aircraft's radar and the line of sight to the target on the ground. When the terrain is flat, the grazing angle equals the depression angle so that they vary directly with each other in size. Hence, in the absence of empirical data regarding the effects of grazing angle on performance, the results of studies of depression angle can be used to formulate general guidelines as to what might be expected. In general, such investigations have revealed that performance is degraded at both very small (less than 10°) and very large (greater than 70°) angles (Spravka, Crawford, & Kuperman, 1990). Whereas the small angles tend to produce long shadows, the large angles often yield poor images because high-reflectance (dihedral) angles become less available to the radar beam as depression angle increases. Thus, performance may improve with increases in angle greater than 10° up to some as yet unknown maximum, where it may then rapidly decline. In the current study, grazing angle ranged from 15° to 45°. Given the outcomes just described, it was expected that performance accuracy would improve as grazing angle increased.

The Theory of Signal Detection

Performance accuracy in a target acquisition task may be evaluated by examining the percentage of correct detections made by the operator. This type of index is problematic, however, because it does not permit an unambiguous interpretation of performance efficiency. For example, the percentage of correct detections may be high because the operator is skilled at differentiating targets from nontargets. Conversely, they may also be high for a poor discriminator who is simply willing to designate almost anything as a target. The theory of signal detection (TSD) is a model of perceptual processing that does provide an estimate of the observer's detection capabilities that is unaffected by his/her general willingness to make a detection response (Gescheider, 1985; Green & Swets, 1966; Macmillan & Creelman, 1991; See, 1994; Wilson, 1992). Consequently, it is frequently used to describe performance outcomes in many types of detection situations, including target acquisition.

Distributions of sensory effects

The fundamental task faced by an observer in a detection situation is to decide on a given trial whether some predefined signal or target did or did not occur. According to TSD, the signal to be detected does not appear in isolation, but rather occurs against a background of noise. Part of the noise is inherent in the sensory process, emanating from the spontaneous firing of the nervous system, while additional noise may arise from changes in the environment or the equipment used to generate the stimuli. Because this noise is always present, the observer's level of sensation will be greater than zero at any moment in time. Consequently, as a result of the presence of noise, the observer's task is not simply to determine whether a signal is present, but instead to decide whether the magnitude of sensation on a given trial is more likely to be the result of noise (N) alone or of signal-plus-noise (SN).

Under the classic parametric model of TSD, the sensory effects produced by N and SN are assumed to follow normal distributions with unit variance, as portrayed in Figure 2. The mean level of excitation produced by the N distribution depends on the background intensity and will be greater than zero since noise is omnipresent. The addition of a signal to the background of noise shifts the level of excitation upward so that the mean of the SN distribution depends on both background and signal intensity. Although the introduction of a signal increases the magnitude of sensation, the nature of the distribution of sensory effects is unchanged (i.e., both the N and SN distributions will be normal with unit variance, differing only in the mean level of excitation).

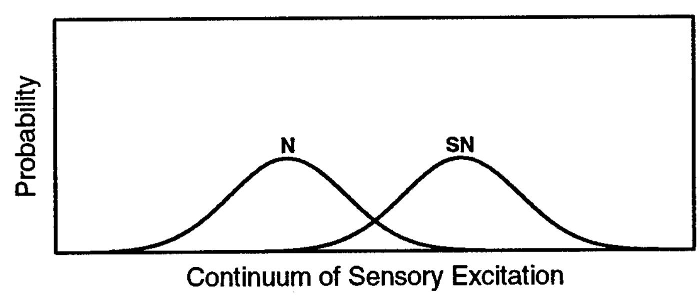

Figure 2. The assumed distribution of sensory effects under TSD.

The likelihood ratio. In deciding whether the magnitude of sensory stimulation is more representative of N or SN, the observer is essentially faced with the task of testing a statistical hypothesis. The observer is assumed under TSD to estimate the probability that an observation arose from SN versus the probability that it arose from N and to compute a “likelihood ratio” of the two probabilities. As the likelihood ratio increases in magnitude, the subjective odds that the stimulus came from the SN distribution become progressively greater. Furthermore, the observer is assumed to establish some critical value so that when the likelihood ratio exceeds this cutoff, the decision is that a signal has been presented. A likelihood ratio that falls below the critical level is assumed to be due to the presence of noise alone.

Decision rules and response outcomes

The observer's decision on any given trial can result in one of four outcomes: (1) a hit (H) occurs if the observer reports "signal" when a signal occurred; (2) a false alarm (FA) results if the observer incorrectly reports "signal" when only the background noise was present; (3) a correct rejection (CR) occurs if "no signal" is reported when only noise was present; and (4) a miss (M) occurs if the observer reports "no signal" when a signal did in fact occur. These outcomes are summarized in Table 1.

Table 1

The Four Possible Decision Outcomes in a Signal Detection Situation

	Stimulus Condition	
Response	**SN (Target)**	**N (Nontarget)**
Signal (Target)	Hit	False Alarm
No Signal (Nontarget)	Miss	Correct Rejection

The critical value or cutoff point set by the observer represents the response criterion, and its location will affect the relative frequency of the four possible outcomes depicted in the table.

The observer who establishes a high cutoff for the likelihood ratio is said to be cautious or conservative since the magnitude of sensation must be very high before the individual will decide that a signal has been presented against the background of noise. Such an observer is less inclined to decide “signal” and more likely to report “no signal” on a given trial. As a result, placement of the criterion upward along the sensory continuum toward the SN distribution results in a decrease in both hits and false alarms as well as an increase in both correct rejections and misses. When the criterion is located midway between the means of the two distributions at their intersection, the observer exhibits no bias toward reporting either “signal” or “no signal.” For this neutral observer, the proportions of hits and correct rejections will be equal as will the proportions of false alarms and misses. On the other hand, when the observer sets a low critical value for the likelihood ratio so that the criterion is shifted downward along the sensory continuum, the response criterion is said to be lenient. This individual requires only a minimal level of sensory excitation to decide that a stimulus came from the SN distribution. Because this observer is more likely to report “signal” than “no signal” on any given trial, both hits and false alarms increase while both correct rejections and misses decrease.

Perceptual sensitivity and response bias

The hit and false alarm responses made by an observer during a detection session are used in the derivation of two independent performance measures: perceptual sensitivity and response bias. Both TSD indices can be calculated using only the proportions of hits and false alarms since the remaining two values are merely their complements: the proportion of misses is equal to 1 - H, and the proportion of correct rejections is equal to 1 - FA. The index of perceptual sensitivity is a perceptual measure that reflects the observer’s ability to discriminate signals from noise, while the response criterion is a nonperceptual index that reflects bias in responding, or the observer's willingness to respond "signal."

The fact that TSD yields two independent measures of detection performance is perhaps its greatest advantage since it permits performance to be characterized independently in terms of both sensing abilities and decision making processes. The two indices are assumed to measure different aspects of performance and to be controlled by different factors. The index of perceptual sensitivity is assumed to be affected only by the sensitivity of the perceptual system to the stimuli for detection, which in turn is affected by such perceptual factors as image resolution and the salience of the signal to be detected. The response criterion, on the other hand, is affected by nonperceptual factors, including the observer's detection goals, expectations about the nature of the stimuli, the probability of signal occurrence, and the anticipated consequences of correct and incorrect responses (payoff). In essence, the application of a detection theory analysis is advantageous because it provides a purer measure of detection ability that is independent of the operator's response bias and is more readily interpreted than separate estimates of hits and false alarms.

Several alternative indices for measuring sensitivity and bias are available under the parametric model of TSD. The traditional measure of perceptual sensitivity, and the one that is most commonly used when the assumptions of normality and equal variance have been met, is the index d' (Green & Swets, 1966; Macmillan & Creelman, 1991). As shown in Figure 3, d' is essentially a measure of the extent of the separation between the means of the SN and N distributions, expressed in terms of the standard deviation of the N distribution:

$$d' = \frac{M_{SN} - M_N}{s_N} \qquad [1]$$

Graphically, the distance between the means of the two distributions grows larger as the observer's sensitivity increases, and the resulting value of d' will increase. The index d' is calculated by determining the z-score that corresponds to the location of the criterion relative to the mean of the SN distribution as well as the z-score that corresponds to its location relative to the mean of the N distribution. The value of d' is given by the following formula:

$$d' = z_{FA} - z_H \qquad [2]$$

According to Craig (1984), d' scores can be used to interpret the level of difficulty of the task with the following "rule-of-thumb" guidelines: very difficult (d' < 1.5), moderately difficult (1.5 to 2.5), moderately easy (2.5 to 3.5), and very easy (d' > 3.5).

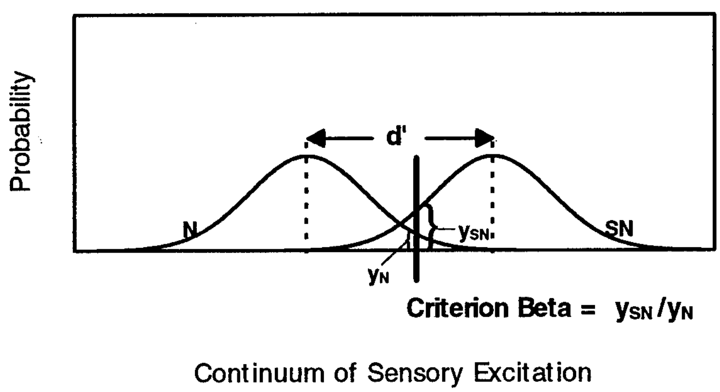

Figure 3. The parametric index of sensitivity, d', is the distance between the means of the N and SN distributions. The parametric index of bias, β, is a ratio of the SN and N ordinates at the criterion.

With regard to parametric measures of response bias, two alternative indices are available: the traditional index, β, and a relatively new measure, c. As portrayed in Figure 3, the value of β can be derived from the point in the sensory continuum where the observer's response criterion is located. More specifically, β is the ratio of the ordinate of the SN distribution to the height of the N distribution at that point:

$$\beta = \frac{\text{ordinate SN at criterion}}{\text{ordinate N at criterion}} \qquad [3]$$

A β of 1.00 signifies a neutral or unbiased observer since the criterion would be placed at a location equidistant from the means of the SN and N distributions where the two ordinates would be identical. In addition, a β greater than 1.00 indicates a conservative criterion, whereas a value between 0.00 and 1.00 represents a lenient criterion.

Whereas the bias index β locates the observer's criterion by the ratio of the ordinates of the SN and N distributions, c locates the criterion by its distance from the intersection of the two distributions measured in z-score units, as depicted in Figure 4. The intersection defines the point where bias is neutral, and location of the criterion at that point yields a c value of 0. Conservative criteria yield positive c values, and liberal criteria produce negative c values. In essence, the index c measures response bias by estimating the extent of the deviation of the observer's criterion from neutrality. The computing formula for c is:

$$c = .5(z_{FA} + z_H) \qquad [4]$$

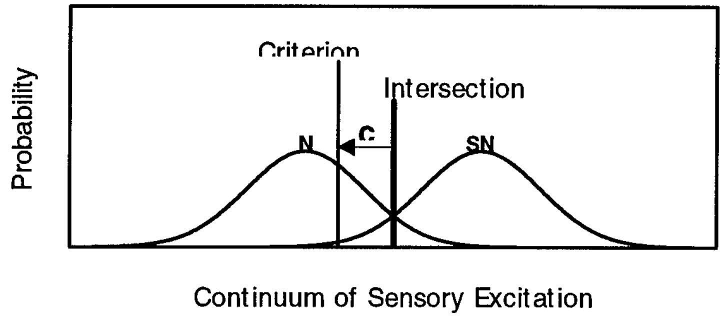

Figure 4. The parametric index of bias, c, is the distance of the criterion from the intersection of the N and SN distributions in z-score units.

A number of studies conducted in the areas of recognition memory and sustained attention have consistently demonstrated that the measure c is superior to β (Macmillan & Creelman, 1991; See, 1994; Snodgrass & Corwin, 1988). The index c is much more responsive than β to nonperceptual manipulations, including signal probability and payoff, and is able to make finer discriminations among conservative and lenient biases (See, 1994). In fact, an in-depth comparison of alternative bias measures in the context of sustained attention revealed that the traditional index, β, is an ineffective measure that is relatively inferior to all other available indices (See, 1994). Consequently, it has been recommended that TSD researchers discontinue using β to represent bias and use the index c instead.

Receiver operating characteristic curves

The performance results of a signal detection task are commonly portrayed through what is referred to as a receiver operating characteristic (ROC) curve. An ROC curve represents the relationship between hit and false alarm probabilities for a given level of sensitivity as the response criterion shifts in a conservative-to-lenient direction. In order to generate an empirical ROC curve, performance data for three or more criterion levels are needed. One approach is to have operators participate in separate sessions wherein variables that affect the response criterion (e.g., signal probability and payoff) are manipulated. However, a more efficient method for generating an ROC curve within a single session is the confidence rating procedure (Gescheider, 1985). In this procedure, operators are asked to supply a confidence rating for each "target" and "nontarget" response, which requires them to use several different response criteria simultaneously.

For example, observers might be asked to rate each response on a scale ranging from 1 (high certainty that a *nontarget* was present) to 6(high certainty that a *target* was present). Since the number of criteria equals the number of confidence rating categories minus one, the operators in this example would be using five criteria simultaneously (C1, C2, C3, C4, and C5). Observations that exceed the most conservative criterion, C5, receive a confidence rating of "six;" those that exceed criterion C4 receive a rating of "five;" and so on. Observations that fall below the most lenient criterion, C1, receive the lowest rating of "one." The proportions of responses for target and nontarget trials for each of the rating categories can be determined and used to derive the hit and false alarm proportions associated with each criterion. The proportions corresponding to the most conservative criterion (C5) are computed from responses that receive a rating of "six." Subsequent hit and false alarm proportions for each of the progressively more lenient criterion levels are cumulative; they are derived by adding the proportions for the appropriate confidence rating level to all preceding proportions. In this example, a five-point ROC curve could be generated from the resulting hit and false alarm proportions corresponding to each of the five criterion levels.

Examples of ROC curves that might be obtained with this technique are depicted in Figure 5. The diagonal line in the figure represents a case in which the observer is unable to discriminate signals from noise (d' = 0). Hence, on the chance diagonal, the proportion of hits equals the proportion of false alarms. For higher levels of sensitivity, the ROC curve is displaced farther from the chance diagonal, as can be seen in the curves in Figure 5 for d' values of .5, 1, 2, and 3. Movement in a left-to-right direction along a single curve represents a conservative-to-lenient change in criterion for that level of sensitivity. ROC curves are particularly useful for determining the trade offs between hits and false alarms that occur both within and between given levels of sensitivity.

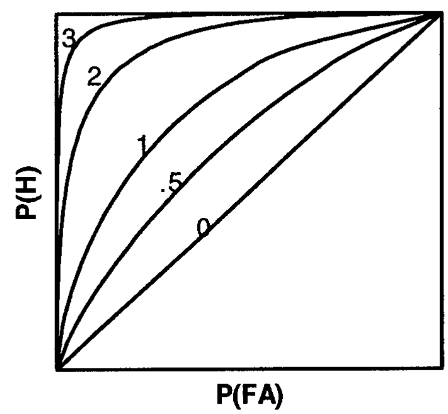

Figure 5. Receiver operating characteristic curves showing d' values of 0, .5, 1, 2, and 3.

Performance Measures in the Present Study

The performance data that were collected in the present study included percentages of hits, false alarms, misses, and correct rejections as well as the reaction times for each type of response. The percentages of hits and false alarms were further used to derive estimates of d' and c. These types of performance data were collected not only because they are the standard dependent variables in target detection tasks but also because they are intended to serve as inputs to ORION, a unique engagement-level effectiveness model currently in use at Armstrong Laboratory (Petersen, Fruchey, Rubin, & O'Rourke, 1995). This model was created to perform mission effectiveness analyses on airborne systems engaged in attacking relocatable, mobile, time critical, or imprecisely located targets. It supports the modeling of multiple, serial sensors and utilizes both correct detections and correct identifications as well as their concomitant false alarm rates. ORION further has the capability to handle several types of target conditions, including targets in the open, in partial obstruction, under netting, amid distracters or decoys, and in various degrees of background clutter. The primary performance input requirements for ORION include estimates of perceptual sensitivity and associated distributions of hits and false alarms as a function of sensor image resolution. Accordingly, the present study was designed to provide these inputs.

METHOD

Subjects

Twelve individuals (two female and ten male) from the military and civilian personnel at Wright-Patterson Air Force Base, OH, volunteered to participate in the study. All subjects met the requirement of unaided or corrected-to-normal 20/20 visual acuity.

Design

Three levels of grazing angle (15°, 30°, and 45°), four levels of image resolution (8.5', 5', 3', and 1'), and three levels of background clutter (low, medium, and high) were combined factorially to provide 36 conditions. Background clutter in this study was defined as the number of trees per square mile. The three levels of clutter used in the study, in order from low to high, were 200, 600, and 1400 trees/mile^2.

Apparatus

Stimuli.

The stimuli were simulated SAR images created via SARSIM with a radar model originally developed at the University of Kansas (Geaga, 1985; Komp, Frost, & Holtzman, 1983). Three types of images were presented in each of the 36 conditions, providing a total of 108 images in the experiment. The three image types included a SCUD, a T-62 tank, and an empty scene devoid of vehicles. The SCUD and tank were "hardbody" reflectors (i.e., there was no signature reduction treatment). Only the SCUD was designated as the target to be detected. The tank and the empty scenes were designated as nontargets. Each image further contained a size cue (graphic inset) that was designed to assist subjects in deciding whether the image was a target or nontarget. The size cue depicted the length of the SCUD target, scaled accordingly for each level of resolution. The set of 108 images was presented in a unique random order for each subject.

Equipment.

The research was conducted in the Visual Image Processing, Enhancement, and Reconstruction (VIPER) facility (Kuperman, Wallquist, & Katz, 1984) located at the Armstrong Laboratory Crew Systems Integration Branch (AL/CFHI) at Wright-Patterson Air Force Base. Images were displayed on an Electrohome Model EVM 1519, 525 line monitor with p-31 (green) phosphor. Prior to the inception of the study, the monitor was calibrated by displaying a 64 step gray scale spanning the full eight bit range of the imagery. The display's gain and contrast controls were adjusted so that each step of the gray scale was discernible and then locked into position for the duration of the experiment. An International Imaging Systems Model 75 image array processor hosted on a Digital Equipment Corporation PDP 11/70 computer system was used to drive the display. The PDP 11/70 also controlled image presentation and data collection.

The apparatus was located in a specially designed light controlled booth where the ambient lighting was held constant at two lux. Subjects responded to each image with a button press on a custom built push-button and trackball control panel, which was interfaced with the PDP 11/70. The trackball was used to bring up each image on the screen, while the push buttons were used to input target/nontarget responses.

Procedure

Upon arrival, subjects were given a brief orientation to the VIPER facility as well as a brief description of the experimental procedures. They were then asked to read the consent form, to ask questions, and to sign the consent form if they wished to participate. Following their consent to participate, subjects were provided with a detailed briefing of the study objectives, the imagery, and task performance requirements.

Subjects participated individually in a single self-paced experimental session, which generally lasted about 20 to 30 minutes. They were seated in a comfortable chair inside the booth at a viewing distance of approximately 71 cm from the display. During each trial, a medium green background with the word READY at the bottom of the screen appeared first. Subjects then initiated a stimulus presentation by moving the trackball on the control panel. A vehicle, if present in the image, always appeared approximately in the center of the screen and always in an open area of the scene so that there was no foliage masking of the vehicle itself. The size cue that accompanied each image was located in the lower right hand corner of the screen. Each image either remained on the screen until a response was input or disappeared after 5 seconds. If a subject had not responded within that time period, the screen blanked and the word RESPOND remained at the bottom of the screen until the subject responded.

Subjects were instructed to respond as quickly and accurately as possible. The confidence rating procedure was employed so that ROC curves might later be generated from the data. Thus, for each stimulus presentation, subjects simultaneously determined if a target was present and indicated their confidence in that decision by pressing one of six labeled buttons on the response panel: (6) target **definitely** present, (5) target **probably** present, (4) target **possibly** present, (3) target **possibly not** present, (2) target **probably not** present, or (1) target **definitely not** present. Thus, if subjects thought the image contained a SCUD target, they pressed either 6, 5, or 4 to represent high, medium, or low certainty, respectively, in their target decision. If they thought the image contained a nontarget tank or no vehicle at all, they pressed either 3, 2, or 1 to signify low, medium, or high certainty, respectively, in their nontarget decision. At the end of each trial, the display screen blanked and the READY prompt for the next trial appeared at the bottom of the screen. The sequence of events for each trial is portrayed in Figure 6.

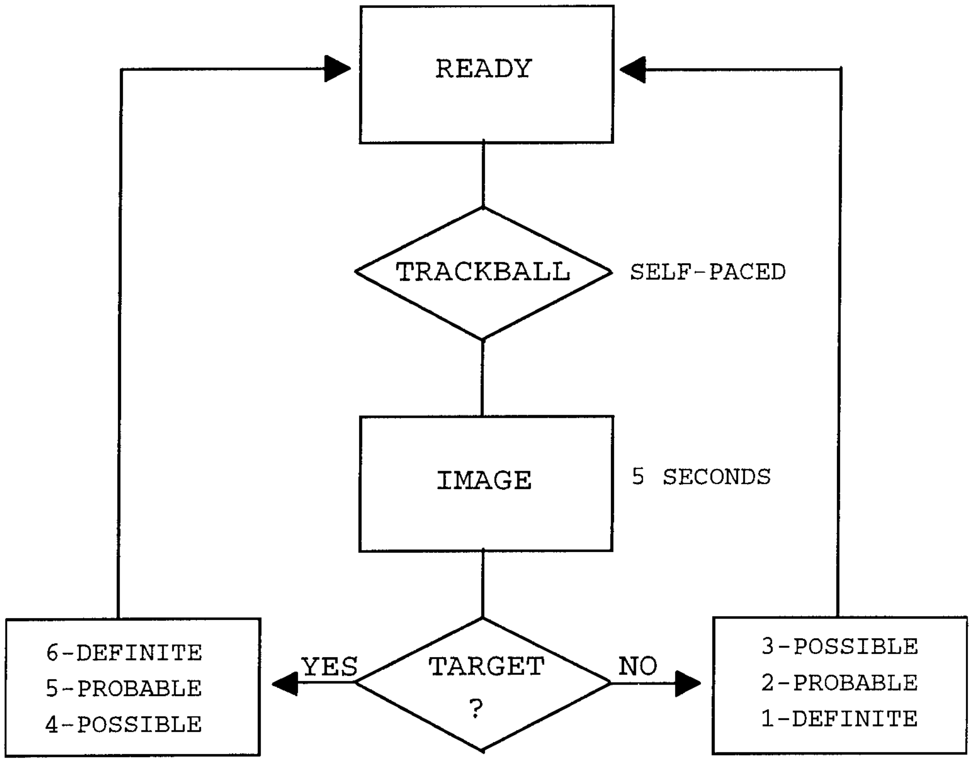

Figure 6. The sequence of events during each trial.

RESULTS

Each subject's responses to each of the 108 images were used to determine the percentages of hits (correct detections of SCUDs), false alarms (incorrect designations of either tanks or empty scenes as targets), misses of targets, and correct rejections of nontargets during the session. As pointed out in the Introduction, since misses and correct rejections are simply the complements of hits and false alarms, respectively, only the latter two dependent variables were analyzed by means of two repeated measures analyses of variance. The probabilities for all F tests were adjusted with the Greenhouse-Geisser epsilon. Any statistically significant main effects were further probed via correlated t tests. The Type I error rate for all post hoc dependent t tests was controlled by means of the Bonferroni procedure, wherein the overall alpha for each set of tests was set at .10.

Hits and False Alarms

Mean percentages of hits and false alarms for each level of angle, resolution, and clutter appear in Table 2. Overall, as can be seen in the table, the mean percentages of correct detections in all conditions were uniformly high, ranging from 85% to 95%. Further, although false alarms did not appear to vary as a function of angle, there was some variation corresponding to the manipulations of resolution and background clutter. With respect to resolution, the percentage of false alarms was similar in all cases except the poorest resolution (8.5'), where the highest percentage of false alarms was observed. With respect to clutter, the percentages of false alarms at medium and high levels of clutter were noticeably higher than at the lowest level of clutter.

Table 2

Means and Standard Deviations for Percentages of Hits and False Alarms in each Condition of Angle, Resolution, and Clutter

CONDITION	PERCENTAGE OF HITS		PERCENTAGE OF FALSE ALARMS	
	Mean	SD	Mean	SD
Angle				
15°	92	16.3	26	10.9
30°	90	17.4	24	14.7
45°	90	15.8	25	11.2
Resolution				
1′	92	22.4	22	18.5
3′	95	16.0	21	17.1
5′	90	16.0	24	16.8
8.5′	85	17.9	35	16.9
Clutter				
Low	92	12.8	19	14.3
Medium	91	21.2	30	14.0
High	90	17.4	27	10.4

A 3(Angle) x 4(Resolution) x 3(Clutter) analysis of variance of the percentage of hits confirmed expectations gained from inspection of Table 2 by revealing no significant main effects or interactions involving angle, resolution, or clutter ($p > .05$). A 3(Angle) x 4(Resolution) x 3(Clutter) analysis of variance of the percentage of false alarms revealed a significant main effect for clutter, $F(2,22) = 7.00$, $p < .014$; a two-way interaction between angle and resolution, $F(6,66)$

= 3.74, $p < .014$; and a three-way interaction between angle, resolution, and clutter, $F(12,132) = 3.62$, $p < .006$. With respect to the main effect for clutter, post hoc correlated t tests indicated that the lowest level of clutter was significantly different from medium and high levels, which themselves did not differ. The analysis of variance further revealed that the apparent differences in mean false alarms as a function of resolution observed in Table 2 were not statistically significant ($p > .05$). Neither the main effect for angle nor any of the remaining two-way interactions in the analysis of false alarms attained statistical significance ($p > .05$).

The nature of the Angle x Resolution interaction for the percentage of false alarms is portrayed in Figure 7. The interaction stems from the differential changes in false alarms that occur for low (8.5' and 5') versus high (3' and 1') image resolutions as grazing angle increases from 15° to 45°. When resolution is high, false alarms remain relatively stable, with the lowest percentage occurring at the moderate grazing angle of 30°. Two very different patterns emerge at the two poorest resolutions. At the 5' resolution, the percentage of false alarms decreases as grazing angle increases; whereas at the 8.5' resolution, the exact opposite occurs.

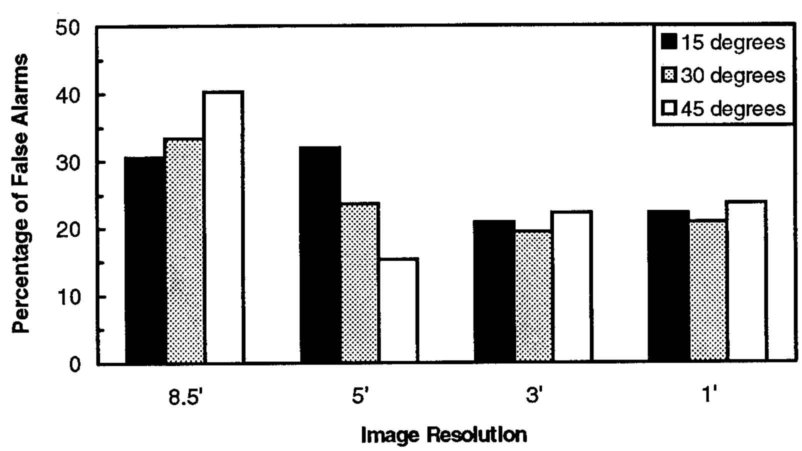

Figure 7. Mean percentage of false alarms for each grazing angle at image resolutions of 8.5', 5', 3', and 1'.

The false alarms were also subjected to a more detailed analysis in order to determine whether most of them were made in response to tanks or to empty scenes. This analysis revealed that subjects committed very few false alarms when the scene was devoid of vehicles. In general, the mean percentage of false alarms to empty scenes was at or near 0%. The most noticeable exception to this trend occurred when image resolution was at its worst at 8.5'. In this condition, the mean percentage of false alarms to empty scenes was 4% ($\underline{SD}$ = 9.9). However, a 3(Angle) x 4(Resolution) x 3(Clutter) analysis of variance revealed that none of the observed differences in means was statistically significant.

Hence, most of the total false alarms that were committed were due to images containing tanks. A 3(Angle) x 4(Resolution) x 3(Clutter) analysis of variance of the percentage of tank false alarms revealed the same pattern of results as the analysis of total false alarms. Namely, the only significant effects were for clutter, the Angle x Resolution interaction, and the three-way interaction.

Perceptual Sensitivity and Response Bias

Each subject's percentages of hits and false alarms were further used to compute estimates of perceptual sensitivity ($\underline{d}'$) and response bias ($\underline{c}$) at each level of angle, resolution, and clutter. Three one-way analyses of variance of $\underline{d}'$ scores and three one-way analyses of $\underline{c}$ scores were conducted to determine whether there were any main effects associated with these three independent variables. As with previous analyses, the probability associated with each $\underline{F}$ test was adjusted with the Greenhouse-Geisser epsilon. Any significant effects were followed up with the correlated $\underline{t}$ test Bonferroni procedure.

Mean d' and c scores for each level of angle, resolution, and clutter appear in Table 3. Overall, the mean sensitivity scores reveal that the target detection task used in this experiment could be considered moderately difficult. The mean scores were similar for various levels of grazing angle but tended to decrease as either resolution became poorer or clutter became more pronounced. As can also be seen in Table 3, mean bias scores were remarkably similar in all conditions. Examination of the mean bias scores in each condition further indicates that subjects were somewhat liberal in their responses (i.e., they were slightly more biased toward making "target" responses as opposed to "nontarget" decisions).

Table 3

Means and Standard Deviations for Perceptual Sensitivity (d') and Response Bias (c) in each Condition of Angle, Resolution, and Clutter

CONDITION	PERCEPTUAL SENSITIVITY (d')		RESPONSE BIAS (c)	
	Mean	SD	Mean	SD
Angle				
15°	2.11	0.61	-0.39	0.35
30°	2.18	0.79	-0.29	0.45
45°	2.07	0.62	-0.34	0.38
Resolution				
1'	2.33	0.73	-0.24	0.61
3'	2.44	0.75	-0.28	0.43
5'	2.11	0.71	-0.22	0.48
8.5'	1.58	0.67	-0.32	0.48

Clutter				
Low	2.40	0.70	-0.21	0.38
Medium	2.03	0.84	-0.42	0.42
High	1.97	0.58	-0.36	0.40

With respect to perceptual sensitivity, the analysis of resolution was statistically significant, $F(3,33) = 7.17$, $p < .003$. Post hoc correlated t tests revealed that perceptual sensitivity at the two highest resolutions of 1' and 3' was significantly better than at the poorest resolution of 8.5'. None of the remaining comparisons was statistically significant. The apparent effect of clutter on sensitivity observed in Table 3 only approached statistical significance, $F(2,22) = 3.94$, $p < .06$, and the analysis of angle was not significant ($p > .05$).

Finally, consistent with observations gained by inspection of Table 3, analyses of mean c scores revealed that subjects' response biases were similar under varying conditions of angle, resolution, and clutter. None of the tests for these effects attained statistical significance ($p > .05$). Practically speaking, this means that subjects were neither more nor less inclined to make a detection response as a function of changes in either angle, image resolution, or background clutter.

Receiver Operating Characteristic Curves

An ROC plot of the four image resolutions used in the study is presented in Figure 8. The five operating characteristics comprising each curve were computed from subjects' proportions of hits and false alarms at each criterion level by means of the procedure described in the Introduction. In ROC space, perceptual sensitivity is portrayed as the distance of each curve from the positive diagonal, where sensitivity is zero. Hence, the ROC plot in Figure 8 shows that sensitivity increased as image resolution improved from 8.5' to 5' to 3' before decreasing slightly at 1'.

The ROC curves further depict the trade offs between hits and false alarms that occur as either sensitivity (image resolution) or bias (confidence rating) varies. For example, when resolution is poor (8.5'), a hit rate of about 80% can be achieved only at the expense of a high false alarm rate near 30%. That is, in order to detect sufficient targets, the observer would have to also be very lenient and incorrectly select many nontarget objects. If sensitivity is enhanced by improving the resolution to 5', a similar hit rate can be achieved with a much lower false alarm rate of about 10%. An operator here would detect the same number of targets with fewer errors of commission. In addition, the ROC plot shows that as response bias becomes more lenient for a fixed level of sensitivity, both hit and false alarm proportions increase. Hence, for a given image resolution, the proportion of false alarms associated with the minimum acceptable level of correct detections can be determined from the ROC curve.

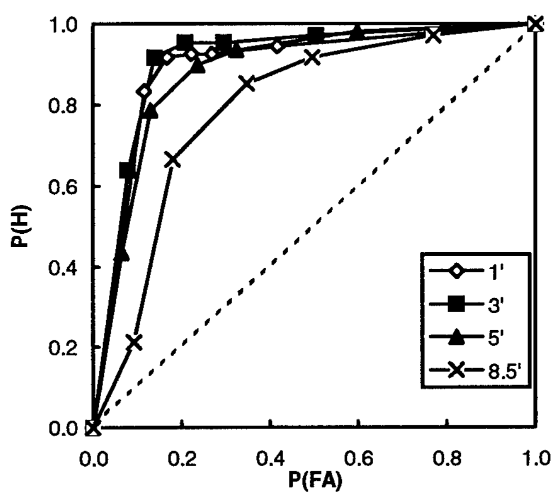

Figure 8. Receiver operating characteristic curves for image resolutions of 8.5', 5', 3', and 1'.

Reaction Time

In addition to collecting performance accuracy data for each subject, the reaction time (RT) in milliseconds for each response was also recorded. Reaction time was defined as the interval between (1) movement of the trackball to initiate an image and (2) the subject's push-button response. Because subjects responded to every image, reaction time was recorded not only for hits and false alarms but also for misses of targets and correct rejections of nontargets. The reaction times for these four types of responses were analyzed by means of a one-way analysis of variance to determine whether there were any differences in the speed of response as a function of response type. Only eight subjects were included in the analysis since four subjects did not miss any targets. The mean reaction times for correct rejections, hits, false alarms, and misses for the eight subjects are plotted in Figure 9. As can be seen in the figure, correct decisions (correct rejections and hits) tended to be made more rapidly than incorrect decisions (false alarms and misses). Despite this trend in sample means, however, the analysis of variance revealed no significant differences among the four types of reaction time, $\underline{F}(3,21) = 2.12$, $\underline{p} > .05$, possibly due to the reduced sample size.

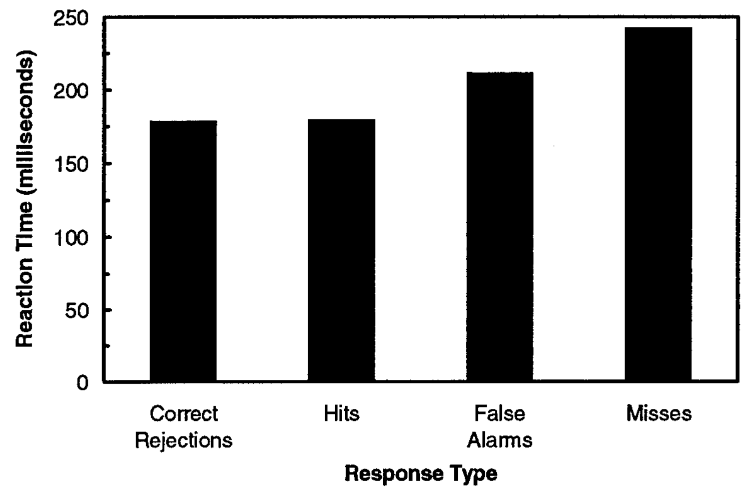

Figure 9. Mean reaction times for correct rejections, hits, false alarms, and misses.

The final set of analyses conducted in this study involved determining whether each type of reaction time was dependent in any way upon angle, resolution, or background clutter. Three separate one-way analyses of variance were conducted for each of the four types of reaction time (correct rejections, hits, false alarms, and misses) for a total of twelve analyses. The only variable that had a significant effect on subjects' reaction times was image resolution. Its effects could be seen in RTs for hits, $F(3,33) = 4.25$, $p < .027$, and for correct rejections of nontargets, $F(3,33) = 17.25$, $p < .0001$. In both instances, as can be seen in Figure 10, RT became progressively faster as image resolution improved from 8.5' to 1'.

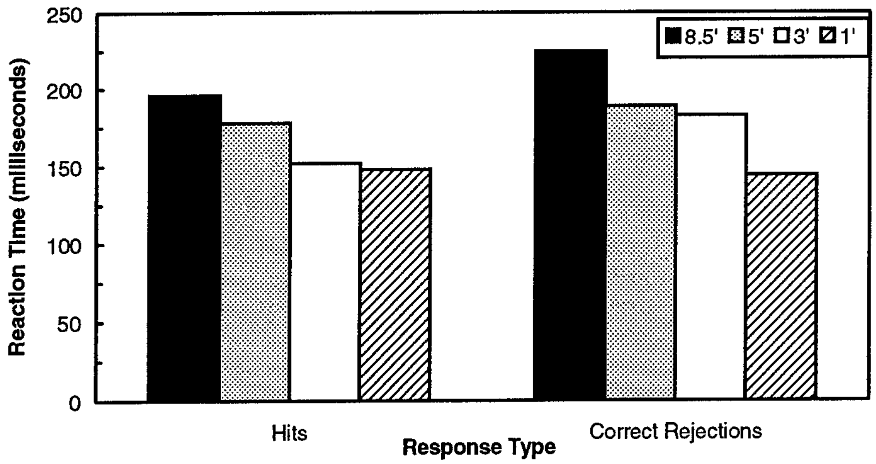

Figure 10. Mean reaction times for hits and correct rejections for image resolutions of 8.5', 5', 3', and 1'.

Post hoc correlated t tests revealed that the mean reaction time for hits was significantly faster at an image resolution of 3' (M = 152, SD = 48.5) versus 8.5' (M = 196, SD = 72). None of the remaining comparisons involving RTs for hits was significant. Follow-up tests for reaction times for correct rejections indicated that all resolutions except 3' and 5' were significantly different from one another. Hence, enhancing image resolution enabled subjects not only to detect targets more quickly but also to correctly reject nontarget objects and empty scenes with alacrity.

CONCLUSIONS

This investigation of the effects of grazing angle, image resolution, and background clutter on performance accuracy, perceptual sensitivity, and reaction time yielded the most consistent outcomes with respect to the manipulation of image resolution. In particular, consistent with previous findings regarding the effects of image resolution on the detection of tactical and relocatable targets (Kuperman, Wilson, & Davis, 1993; Kuperman, Wilson, & Perez, 1988), improvements in image resolution enhanced the detection of TM-type targets in the present study up to the point of the 3' resolution, whereupon further increments in resolution no longer enhanced performance. However, these effects could not be observed merely by examining isolated hit and false alarm scores; rather, they emerged only through an analysis of perceptual sensitivity, a measure of detection efficiency which simultaneously takes both hits and false alarms into consideration. Further, subjects were not only more sensitive at the higher resolutions but also faster at detecting target objects and correctly rejecting nontarget images. In this study, the sole indication that a 1' resolution might represent an improvement over the 3' resolution was the finding that subjects were significantly faster at correctly rejecting nontarget images at the finest resolution; sensitivity to target detection itself did not differ as resolution improved from 3' to 1'.

Outcomes involving the other two independent variables revealed that detection efficiency per se was not affected by either grazing angle or background clutter. However, grazing angle did have an impact on errors of commission, the nature of which varied depending upon the level of image resolution that was present. The absence of an association in the present study between detection efficiency and variations in grazing angle ranging from 15° to 45° suggests that the maximum grazing angle for optimal detection performance may lie somewhere between 45° and

70°. The effects of clutter also emerged only in the analysis of errors of commission, with false alarms being significantly greater at medium and high clutter levels. Hence, as clutter becomes more pronounced, operators may be more likely to incorrectly designate a nontarget object as a target. It should be pointed out that these outcomes pertain to clutter when defined as the number of trees per square mile present in an image. Alternative definitions of clutter may yield different results (see appendix).

A fine-grained analysis of subjects' errors of commission revealed that the majority of false alarms occurred with tanks as opposed to empty scenes. Further, although false alarms to empty scenes were rare, they reached their peak when image resolution was poorest (8.5'). These findings imply that an operator will generally commit a false alarm only when a distracter object is present, unless image quality is so poor as to make it difficult to differentiate between clutter and potential objects of interest.

Finally, the results of the ROC analysis in the present study provide the required inputs for the ORION engagement-level effectiveness model. The obtained ROC curves can be used to derive estimates of sensitivity for the four levels of image resolution included in the study. They can also be used to obtain expected distributions of hits and false alarms as a function of sensitivity for each level of resolution. It should be noted that these and other results reported here most likely represent the *maximum* level of operator performance that can be expected under the conditions of angle, resolution, and clutter that were examined in the study, given that potential targets were presented in the absence of both foliage masking and signature reduction treatment.

REFERENCES

Cathcart, J. M., Doll, T. J., & Schmieder, D. E. (1989). Target detection in urban clutter. IEEE Transactions on Systems, Man, and Cybernetics, 19, 1242-1250.

Craig, A. (1984). Human engineering: The control of vigilance. In J. S. Warm (Ed.), Sustained attention in human performance (pp. 247-291). Chichester, U.K.: Wiley.

Geaga, J. V. (1985). Synthetic aperture radar target and terrain simulator (Report No. 86-4R). Palos Verdes Peninsula, CA: Autonomous Systems Laboratory, Northrop Research and Technology Center.

Gescheider, G. A. (1985). Psychophysics: Method, theory, and application (2nd ed.). Hillsdale, NJ: Erlbaum.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley.

Komp, E. E., Frost, V. S., & Holtzman, J. C. (1983). User manual for the radar image simulator (Report No. RSL TR 581-1). Lawrence, KS: Remote Sensing Laboratory, University of Kansas Center for Research.

Kuperman, G. G., Wallquist, D. L., & Katz, L. (1984). A digital image processing facility for human factors research. Proceedings of the Human Factors Society 28th Annual Meeting. San Antonio, TX.

Kuperman, G. G., Wilson, D. L., & Davis, I. (1993). High resolution radar demonstration program: Operator performance study (Report No. AL-TR-1993-0032). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

Kuperman, G. G., Wilson, D. L., & Perez, W. A. (1988). Relocatable target acquisition performance with simulated synthetic aperture radar imagery (Report No. AAMRL-TR-88-025). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. Cambridge: Cambridge University Press.

Petersen, J., Fruchey, D., Rubin, R., & O'Rourke, P. (1995). ORION sortie effectiveness model: User's manual. Pico Rivera, CA: Northrop Grumman Corporation.

See, J. E. (1994). Vigilance and signal detection theory: An evaluation of response bias measures. Unpublished doctoral dissertation, University of Cincinnati, Cincinnati, OH.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. Journal of Experimental Psychology: General, 117, 34-50.

Spravka, J. J., Crawford, R. L., & Kuperman, G. G. (1990). A human factors review of the synthetic aperture radar literature (Report No. AAMRL-TR-90-022). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Toms, M. L., & Kuperman, G. G. (1991). Sensor fusion: A human factors perspective (Report No. AL-TR-1991-0152). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

Wilson, D. L. (1992). Theory of signal detection and its application to visual target acquisition: A review of the literature (Report No. AL-TR-1992-0083). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

APPENDIX

An Alternative Representation of Clutter

In the present study, background clutter was defined as the number of trees per square mile present in an image. Given that an increase in foliage can mask potential target objects and make them more difficult to detect, one might expect decrements in detection efficiency to occur as the number of trees per square mile increases. In addition, one might also expect longer scan times and thus longer reaction times in such situations. However, none of these effects were observed in the current investigation. The sole outcome associated with clutter was a tendency for operators to commit more false alarms under medium and high as opposed to low clutter.

In an effort to explore the validity of designating clutter as the number of trees per square mile, the relationship between this clutter metric and an alternative metric that has been used successfully to represent the magnitude of clutter in rural scenes was examined (Cathcart, Doll, & Schmieder, 1989). To compute the latter metric, an image must first be divided into cells whose size corresponds to the size of the target to be detected. Next, the radiance standard deviation (σ_i) in each cell is calculated:

$$\sigma_i = \left(\frac{\sum_{j=1}^{k} \left(X_{ij} - \mu \right)^2}{k} \right)^{1/2} \qquad [5]$$

In the formula, $\underline{X}_{ij}$ represents the intensity of the $\underline{j}$th pixel in the $\underline{i}$th cell of the image; μ is the average intensity in the $\underline{i}$th cell; and $\underline{k}$ is the total number of pixels in the cell.

The average of the radiance standard deviations over the total number of cells (N) in the entire image can then be computed with the following formula:

$$clutter = \frac{\left|\sum \sigma_i^2\right|^{\frac{1}{2}}}{N} \qquad [6]$$

In effect, this clutter metric provides an estimate of the average deviation in intensities within an image. The rationale behind this measure is that the average deviation should increase as background clutter increases and introduces more and more variations in intensity.

The procedure used by Cathcart, Doll, and Schmieder (1989) was applied in the current study in order to obtain an alternative clutter metric for each of the 108 images. The size of the SCUD target as a function of resolution was used to define the cell size N in Equation 6. The estimates of clutter derived from the application of this procedure were then analyzed to determine if they bore any sort of relationship to our previous designation of clutter as the number of trees per square mile.

The mean computed estimates of clutter for our low, medium, and high clutter levels were 1.5 ($SD = .65$), 1.3 ($SD = .70$), and 1.6 ($SD = .70$). Simple observation of these means indicates that the computed metric changes very little as clutter defined in the present study shifts from low to high. This observation was confirmed statistically by computing a partial correlation between the two estimates that controlled for the angle, resolution, and type of object (none, tank, or SCUD) present in each of the 108 images. The resulting correlation was $r = .16$, $p > .10$. Thus, across the set of images, less than 3% of the variation in one clutter metric corresponded to variation in the second metric, which signifies that the two "clutter" metrics are not measuring similar aspects of a given image.

One reason for the absence of a relationship may have been insufficient variability in intensity values within each image. The fact that trees were the only type of clutter that could appear in each image may have artificially imposed limits on the range of intensity values that could occur, which in turn would limit the range of possible values for the computed clutter metric (i.e., an increment in the number of trees may represent an increase in the number of distracting/masking objects but not an increase in variations in intensity throughout the image).

While the computed clutter metric proved to be unrelated to our designation of clutter as the number of trees per square mile, the possibility existed that the new metric might bear some relationship to the performance data that were collected in the present study. In order to explore this possibility, correlations between the new clutter metric and hits, false alarms, and reaction times for hits and correct rejections were examined. These dependent variables were selected in order to avoid the reduction in sample size and power that can occur when a variable that has missing data is used (e.g., RT for misses when there are some subjects who did not miss any targets). The first set of analyses was conducted for the target images used in the present study. Across the 12 subjects participating in the experiment, the mean percentage of hits and mean RT for hits associated with the 36 SCUD target images were obtained. Partial correlations controlling for angle and resolution were then computed between each of these dependent variables and the new clutter metric. The correlations revealed that higher clutter levels were associated with lower percentages of hits ($\underline{r} = -.18$, $\underline{p} > .30$) and longer reaction times for hits ($\underline{r}$ =

.35, $p < .045$). However, while the direction of each relationship was in the expected direction, only the correlation between RT for hits and clutter attained statistical significance.

The second set of analyses was conducted for the nontarget images that were used in the current investigation. The mean percentage of false alarms and the mean RT for correct rejections of nontargets were obtained for the 72 tank and empty images. Partial correlations controlling for angle, resolution, and type of object (none or tank) were then computed between each of these dependent variables and the new clutter metric. The resulting correlations indicated that higher levels of clutter were associated with higher false alarm rates, though the relationship was not statistically significant ($r = .12$, $p > .34$). RT for CRs was unrelated to clutter ($r = -.07$, $p > .56$).

Overall, the correlational analyses reported in this appendix suggest that the new clutter metric, though unrelated to the manner in which clutter had previously been defined in the current study, does bear some relationship to standard performance metrics. According to Cathcart, Doll, and Schmieder (1989), however, its utility may be limited to defining the magnitude of the type of ambiguous and unpredictable clutter characteristic of rural scenes (e.g., brush, trees, and bodies of water) as opposed to the more contextual structure provided by urban clutter (e.g., buildings, roads, and vehicles).

In addition, these analyses serve to illustrate that the single label "clutter" has multiple and possibly unrelated connotations. As a result, outcomes regarding the quality of detection performance as a function of clutter magnitude will almost certainly be highly dependent upon the precise definition of clutter used in a given task and may not generalize to situations in which clutter is designated differently. The analyses further emphasize the continued need to search for a more universally acceptable definition of clutter that can be used for various background types (e.g., urban versus rural), as stated in an earlier report by Cathcart, Doll, and Schmieder (1989). As they point out, a good clutter definition is one that is capable of quantifying the essence of those background features that *interfere* with target detection.

GLOSSARY

ACC	Air Combat Command
AO	Attack Operations
ATC	Automatic Target Cueing
ATR	Automatic Target Recognition
β	Response bias index
$\underline{c}$	Response bias index
CAD	Computer-Aided Design
CD&E	Concept Definition and Exploration
CR	Correct rejection
$\underline{d}'$	Perceptual sensitivity
FA	False alarm
H	Hit
IA	Image Analyst
M	Miss
N	Noise
ROC	Receiver Operating Characteristic
RT	Reaction Time
SAR	Synthetic Aperture Radar
SN	Signal-plus-noise
TEL	Transporter/Erector/Launcher
TM	Theater Missile
TMD	Theater Missile Defense
TSD	Theory of Signal Detection
UAV	Unmanned Air Vehicle
VIPER	Visual Image Processing, Enhancement, and Reconstruction
WSO	Weapon Systems Officer

U. S. Government Printing Office 1996 750-071/00173